The Unicode® Standard
Version 15.0 – Core Specification

To learn about the latest version of the Unicode Standard, see https://www.unicode.org/versions/latest/.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.

The authors and publisher have taken care in the preparation of this specification, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The *Unicode Character Database* and other files are provided as-is by Unicode, Inc. No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided.

© 2022 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction. For information regarding permissions, inquire at https://www.unicode.org/reporting.html. For information about the Unicode terms of use, please see https://www.unicode.org/copyright.html.

The Unicode Standard / the Unicode Consortium; edited by the Unicode Consortium. — Version 15.0.

Includes index.

ISBN 978-1-936213-32-0 (https://www.unicode.org/versions/Unicode15.0.0/)

1. Unicode (Computer character set) I. Unicode Consortium.

QA268.U545 2022

ISBN 978-1-936213-32-0

Published in Mountain View, CA

September 2022

Contents

Preface xxi
Why Unicode? xxi
Organization of This Standard xxi
The Unicode Character Database xxii
Unicode Code Charts xxiii
Unicode Standard Annexes xxiii
Unicode Technical Standards and Unicode Technical Reports xxiv
Updates and Errata xxv
Acknowledgements xxv
1 Introduction 1
1.1 Coverage 3
Standards Coverage 3
New Characters 3
1.2 Design Goals 4
1.3 Text Handling 6
Characters and Glyphs 6
Text Elements 6
2 General Structure 9
2.1 Architectural Context 10
Basic Text Processes 10
Text Elements, Characters, and Text Processes 10
Text Processes and Encoding 11
2.2 Unicode Design Principles 14
Universality 14
Efficiency 15
Characters, Not Glyphs 15
Semantics 18
Plain Text 18
Logical Order 19
Unification 21
Dynamic Composition 23
Stability 23
Convertibility 24
2.3 Compatibility Characters 25
Compatibility Variants 26
Compatibility Decomposable Characters 26
2.4 Code Points and Characters 29
Types of Code Points 30

2.5 Encoding Forms ... 33
UTF-32 ... 35
UTF-16 ... 36
UTF-8 ... 37
2.6 Encoding Schemes ... 39
2.7 Unicode Strings ... 42
2.8 Unicode Allocation ... 43
Planes ... 43
Allocation Areas and Blocks ... 44
Assignment of Code Points ... 45
2.9 Details of Allocation ... 47
Plane 0 (BMP) ... 48
Plane 1 (SMP) ... 50
Plane 2 (SIP) ... 51
Plane 3 (TIP) ... 51
Other Planes ... 51
2.10 Writing Direction ... 52
2.11 Combining Characters ... 54
Sequence of Base Characters and Combining Characters ... 55
Multiple Combining Characters ... 56
Ligated Multiple Base Characters ... 59
Exhibiting Nonspacing Marks in Isolation ... 60
"Characters" and Grapheme Clusters ... 60
2.12 Equivalent Sequences ... 61
Normalization ... 61
Decompositions ... 62
Non-decomposition of Certain Diacritics ... 64
2.13 Special Characters ... 66
Special Noncharacter Code Points ... 66
Byte Order Mark (BOM) ... 66
Layout and Format Control Characters ... 67
The Replacement Character ... 67
Control Codes ... 67
2.14 Conforming to the Unicode Standard ... 68
Characteristics of Conformant Implementations ... 68
Unacceptable Behavior ... 69
Acceptable Behavior ... 69
Supported Subsets ... 70
3 Conformance ... 71
3.1 Versions of the Unicode Standard ... 72
Stability ... 72
Version Numbering ... 73
Errata and Corrigenda ... 74

References to the Unicode Standard 74
Precision in Version Citation 75
References to Unicode Character Properties 75
References to Unicode Algorithms 76
3.2 Conformance Requirements 77
Code Points Unassigned to Abstract Characters 77
Interpretation 78
Modification 79
Character Encoding Forms 80
Character Encoding Schemes 81
Bidirectional Text 82
Normalization Forms 82
Normative References 82
Unicode Algorithms 82
Default Casing Algorithms 83
Unicode Standard Annexes 83
3.3 Semantics 85
Definitions 85
Character Identity and Semantics 85
3.4 Characters and Encoding 88
3.5 Properties 93
Types of Properties 93
Property Values 94
Default Property Values 95
Classification of Properties by Their Values 96
Property Status 97
Context Dependence 100
Stability of Properties 101
Simple and Derived Properties 102
Property Aliases 103
Private Use 104
3.6 Combination 105
Combining Character Sequences 105
Grapheme Clusters 108
Application of Combining Marks 109
3.7 Decomposition 115
Compatibility Decomposition 115
Canonical Decomposition 116
3.8 Surrogates 118
3.9 Unicode Encoding Forms 119
UTF-32 123
UTF-16 124
UTF-8 124

Encoding Form Conversion 126
Constraints on Conversion Processes 126
U+FFFD Substitution of Maximal Subparts 127
3.10 Unicode Encoding Schemes 130
3.11 Normalization Forms 134
Normalization Stability 134
Combining Classes 135
Specification of Unicode Normalization Forms 136
Starters 136
Canonical Ordering Algorithm 137
Canonical Composition Algorithm 138
Definition of Normalization Forms 140
3.12 Conjoining Jamo Behavior 142
Definitions 142
Hangul Syllable Decomposition 144
Hangul Syllable Composition 146
Hangul Syllable Name Generation 147
Sample Code for Hangul Algorithms 148
3.13 Default Case Algorithms 152
Definitions 152
Default Case Conversion 154
Default Case Folding 155
Default Case Detection 156
Default Caseless Matching 157
4 Character Properties 159
4.1 Unicode Character Database 161
4.2 Case 164
Definitions of Case and Casing 164
Case Mapping 166
4.3 Combining Classes 168
Reordrant, Split, and Subjoined Combining Marks 168
4.4 Directionality 171
4.5 General Category 172
4.6 Numeric Value 175
Ideographic Numeric Values 176
4.7 Bidi Mirrored 178
4.8 Name 180
Unicode Name Property 183
Code Point Labels 186
Use of Character Names in APIs and User Interfaces 186
4.9 Unicode 1.0 Names 188
4.10 Letters, Alphabetic, and Ideographic 189

4.11 Properties for Text Boundaries 190
4.12 Characters with Unusual Properties 191
5 Implementation Guidelines 195
5.1 Data Structures for Character Conversion 196
Issues 196
Multistage Tables 196
5.2 Programming Languages and Data Types 199
Unicode Data Types for C 199
5.3 Unknown and Missing Characters 201
5.4 Handling Surrogate Pairs in UTF-16 203
5.5 Handling Numbers 205
5.6 Normalization 206
5.7 Compression 208
5.8 Newline Guidelines 209
Definitions 209
Line Separator and Paragraph Separator 210
Recommendations 211
5.9 Regular Expressions 214
5.10 Language Information in Plain Text 215
Requirements for Language Tagging 215
Language Tags and Han Unification 215
5.11 Editing and Selection 217
Consistent Text Elements 217
5.12 Strategies for Handling Nonspacing Marks 219
Keyboard Input 220
Truncation 220
5.13 Rendering Nonspacing Marks 222
Canonical Equivalence 225
Positioning Methods 226
5.14 Locating Text Element Boundaries 228
5.15 Identifiers 229
5.16 Sorting and Searching 230
Culturally Expected Sorting and Searching 230
Language-Insensitive Sorting 230
Searching 231
Sublinear Searching 232
5.17 Binary Order 233
UTF-8 in UTF-16 Order 233
UTF-16 in UTF-8 Order 234

5.18 Case Mappings 236
Titlecasing 236
Complications for Case Mapping 237
Reversibility 239
Caseless Matching 240
Normalization and Casing 242
5.19 Mapping Compatibility Variants 243
5.20 Unicode Security 245
5.21 Ignoring Characters in Processing 248
Characters Ignored in Text Segmentation 248
Characters Ignored in Line Breaking 249
Characters Ignored in Cursive Joining 249
Characters Ignored in Identifiers 249
Characters Ignored in Searching and Sorting 250
Characters Ignored for Display 250
5.22 U+FFFD Substitution in Conversion 255
6 Writing Systems and Punctuation 257
6.1 Writing Systems 258
6.2 General Punctuation 264
Blocks Devoted to Punctuation 265
Format Control Characters 266
Space Characters 267
Dashes and Hyphens 268
Paired Punctuation 270
Language-Based Usage of Quotation Marks 271
Apostrophes 275
Other Punctuation 276
Archaic Punctuation and Editorial Marks 281
Indic Punctuation 283
CJK Punctuation 286
Unknown or Unavailable Ideographs 287
CJK Compatibility Forms 288
7 Europe-I 289
Modern and Liturgical Scripts
7.1 Latin 291
Letters of Basic Latin: U+0041–U+007A 295
Letters of the Latin-1 Supplement: U+00C0–U+00FF 295
Latin Extended-A: U+0100–U+017F 295
Latin Extended-B: U+0180–U+024F 296
IPA Extensions: U+0250–U+02AF 298
Phonetic Extensions: U+1D00–U+1DBF 300
Latin Extended Additional: U+1E00–U+1EFF 301
Latin Extended-C: U+2C60–U+2C7F 301

Latin Extended-D: U+A720–U+A7FF 302
Latin Extended-E: U+AB30–U+AB6F 303
Latin Extended-F: U+10780–U+107BF 304
Latin Ligatures: U+FB00–U+FB06 304
7.2 Greek 305
Greek: U+0370–U+03FF 305
Greek Extended: U+1F00–U+1FFF 309
Ancient Greek Numbers: U+10140–U+1018F 310
7.3 Coptic 312
7.4 Cyrillic 315
Cyrillic: U+0400–U+04FF 315
Cyrillic Supplement: U+0500–U+052F 317
Cyrillic Extended-A: U+2DE0–U+2DFF 317
Cyrillic Extended-B: U+A640–U+A69F 318
Cyrillic Extended-C: U+1C80–U+1C8F 318
Cyrillic Extended-D: U+1E030–U+1E08F 319
7.5 Glagolitic 320
Glagolitic: U+2C00–U+2C5F 320
Glagolitic Supplement: U+1E000–U+1E02F 320
7.6 Armenian 321
7.7 Georgian 323
Georgian: U+10A0–U+10FF 323
Georgian Extended: U+1C90–U+1CBF 323
Georgian Supplement: U+2D00–U+2D2F 323
7.8 Modifier Letters 325
Spacing Modifier Letters: U+02B0–U+02FF 326
Modifier Tone Letters: U+A700–U+A71F 328
7.9 Combining Marks 329
Combining Diacritical Marks: U+0300–U+036F 334
Combining Diacritical Marks Extended: U+1AB0–U+1AFF 335
Combining Diacritical Marks Supplement: U+1DC0–U+1DFF 336
Combining Diacritical Marks for Symbols: U+20D0–U+20FF 336
Combining Half Marks: U+FE20–U+FE2F 337
Combining Marks in Other Blocks 338
8 Europe-II 341
Ancient and Other Scripts
8.1 Linear A 343
8.2 Linear B 344
Linear B Syllabary: U+10000–U+1007F 344
Linear B Ideograms: U+10080–U+100FF 344
Aegean Numbers: U+10100–U+1013F 344
8.3 Cypriot Syllabary 346

8.4 Cypro-Minoan 347
8.5 Ancient Anatolian Alphabets 348
Lycian: U+10280–U+1029F 348
Carian: U+102A0–U+102DF 348
Lydian: U+10920–U+1093F 348
8.6 Old Italic 350
8.7 Runic 353
8.8 Old Hungarian 356
8.9 Gothic 357
8.10 Elbasan 358
8.11 Caucasian Albanian 359
8.12 Vithkuqi 360
8.13 Old Permic 361
8.14 Ogham 362
8.15 Shavian 363
9 Middle East-I 365
Modern and Liturgical Scripts
9.1 Hebrew 367
Hebrew: U+0590–U+05FF 367
Alphabetic Presentation Forms: U+FB1D–U+FB4F 371
9.2 Arabic 373
Arabic: U+0600–U+06FF 373
Arabic Cursive Joining 382
Arabic Ligatures 385
Arabic Joining Groups 387
Combining Hamza 395
Other Letters for Extended Arabic 397
Arabic Letters in Other Languages 397
Arabic Supplement: U+0750–U+077F 398
Arabic Extended-A: U+08A0–U+08FF 398
Arabic Presentation Forms-A: U+FB50–U+FDFF 398
Arabic Presentation Forms-B: U+FE70–U+FEFF 400
9.3 Syriac 401
Syriac: U+0700–U+074F 401
Syriac Shaping 405
Syriac Supplement: U+0860–U+086F 410
9.4 Samaritan 411
9.5 Mandaic 413
9.6 Yezidi 416

10 Middle East-II 417
Ancient Scripts
10.1 Old North Arabian 419
10.2 Old South Arabian 420
10.3 Phoenician 422
10.4 Imperial Aramaic 424
10.5 Manichaean 426
10.6 Pahlavi and Parthian 430
Inscriptional Parthian: U+10B40–U+10B5F 430
Inscriptional Pahlavi: U+10B60–U+10B7F 430
Psalter Pahlavi: U+10B80–U+10BAF 431
10.7 Avestan 432
10.8 Chorasmian 434
10.9 Elymaic 435
10.10 Nabataean 436
10.11 Palmyrene 437
10.12 Hatran 438
11 Cuneiform and Hieroglyphs 439
11.1 Sumero-Akkadian 440
Cuneiform: U+12000–U+123FF 440
Cuneiform Numbers and Punctuation: U+12400–U+1247F 443
Early Dynastic Cuneiform: U+12480–U+1254F 443
11.2 Ugaritic 444
11.3 Old Persian 445
11.4 Egyptian Hieroglyphs 446
Egyptian Hieroglyphs: U+13000–U+1342F 446
Egyptian Hieroglyph Format Controls: U+13430–U+1345F 448
Editorial Marks 453
11.5 Meroitic 456
11.6 Anatolian Hieroglyphs 458
12 South and Central Asia-I 461
Official Scripts of India
12.1 Devanagari 463
Devanagari: U+0900–U+097F 463
Principles of the Devanagari Script 464
Rendering Devanagari 473
Devanagari Digits, Punctuation, and Symbols 483
Extensions in the Main Devanagari Block 483
Devanagari Extended: U+A8E0–U+A8FF 485
Devanagari Extended-A: U+11B00–U+11B5F 487
Vedic Extensions: U+1CD0–U+1CFF 487

12.2 Bengali (Bangla) 489
12.3 Gurmukhi 496
12.4 Gujarati 501
12.5 Oriya (Odia) 503
12.6 Tamil 507
Tamil: U+0B80–U+0BFF 507
Tamil Vowels 508
Tamil Ligatures 510
Tamil Supplement: U+11FC0–U+11FFF 514
Tamil Named Character Sequences 514
12.7 Telugu 517
12.8 Kannada 520
Kannada: U+0C80–U+0CFF 520
Principles of the Kannada Script 520
Rendering Kannada 522
12.9 Malayalam 524
Malayalam: U+0D00–U+0D7F 524
Malayalam Orthographic Reform 525
Rendering Malayalam 526
Malayalam Numbers and Punctuation 531
13 South and Central Asia-II 533
Other Modern Scripts
13.1 Thaana 535
13.2 Sinhala 537
Sinhala: U+0D80–U+0DFF 537
Sinhala Archaic Numbers: U+111E0–U+111FF 540
13.3 Newa 541
13.4 Tibetan 544
13.5 Mongolian 555
Mongolian: U+1800–U+18AF 555
Mongolian Supplement: U+11660–U+1167F 564
13.6 Limbu 565
13.7 Meetei Mayek 569
Meetei Mayek: U+ABC0–U+ABFF 569
Meetei Mayek Extensions: U+AAE0–U+AAF6 570
13.8 Mro 571
13.9 Warang Citi 572
13.10 Ol Chiki 573
13.11 Chakma 575
13.12 Lepcha 576
13.13 Saurashtra 579

13.14 Masaram Gondi 580
13.15 Gunjala Gondi 582
13.16 Wancho 584
13.17 Toto 585
13.18 Tangsa 586
14 South and Central Asia-III 587
Ancient Scripts
14.1 Brahmi 589
14.2 Kharoshthi 593
Kharoshthi: U+10A00–U+10A5F 593
Rendering Kharoshthi 595
14.3 Bhaiksuki 599
14.4 Phags-pa 601
14.5 Marchen 608
14.6 Zanabazar Square 609
14.7 Soyombo 612
14.8 Old Turkic 614
14.9 Old Sogdian 615
14.10 Sogdian 616
14.11 Old Uyghur 617
15 South and Central Asia-IV 619
Other Historic Scripts
15.1 Syloti Nagri 622
15.2 Kaithi 624
15.3 Sharada 627
15.4 Takri 629
15.5 Siddham 631
15.6 Mahajani 633
15.7 Khojki 635
15.8 Nag Mundari 637
15.9 Khudawadi 638
15.10 Multani 640
15.11 Tirhuta 641
15.12 Modi 644
15.13 Nandinagari 647
Nandinagari: U+119A0–U+119FF 647
15.14 Grantha 649
Grantha: U+11300–U+1137F 649
Rendering Grantha 649

15.15 Dives Akuru ... 653
15.16 Ahom ... 655
15.17 Sora Sompeng ... 657
15.18 Dogra ... 658
16 Southeast Asia-I ... 661
Thailand, Laos, Myanmar, Cambodia, Vietnam
16.1 Thai ... 663
16.2 Lao ... 667
16.3 Myanmar ... 670
Myanmar: U+1000–U+109F ... 670
Myanmar Extended-A: U+AA60–U+AA7F ... 674
Khamti Shan ... 674
Aiton and Phake ... 676
Myanmar Extended-B: U+A9E0–U+A9FF ... 676
16.4 Khmer ... 677
Khmer: U+1780–U+17FF ... 677
Principles of the Khmer Script ... 677
Khmer Symbols: U+19E0–U+19FF ... 688
16.5 Tai Le ... 689
16.6 New Tai Lue ... 691
16.7 Tai Tham ... 694
16.8 Tai Viet ... 697
16.9 Kayah Li ... 700
16.10 Cham ... 701
16.11 Pahawh Hmong ... 703
16.12 Nyiakeng Puachue Hmong ... 705
16.13 Pau Cin Hau ... 707
16.14 Hanifi Rohingya ... 708
17 Southeast Asia-II ... 709
Indonesia and the Philippines
17.1 Philippine Scripts ... 710
Tagalog: U+1700–U+171F ... 710
Hanunóo: U+1720–U+173F ... 710
Buhid: U+1740–U+175F ... 710
Tagbanwa: U+1760–U+177F ... 710
Principles of the Philippine Scripts ... 710
17.2 Buginese ... 713
17.3 Balinese ... 715
17.4 Javanese ... 721
17.5 Rejang ... 724

17.6 Batak 725
17.7 Sundanese 727
Sundanese: U+1B80–U+1BBF 727
Sundanese Supplement: U+1CC0–U+1CCF 729
17.8 Makasar 730
17.9 Kawi 732
18 East Asia 737
18.1 Han 740
CJK Unified Ideographs 740
Blocks Containing Han Ideographs 741
General Characteristics of Han Ideographs 744
Principles of Han Unification 748
Unification Rules 749
Abstract Shape 750
Han Ideograph Arrangement 752
Radical-Stroke Indices 753
Mappings for Han Ideographs 754
CJK Compatibility Ideographs: U+F900–U+FAFF 755
CJK Compatibility Supplement: U+2F800–U+2FA1D 756
Kanbun: U+3190–U+319F 756
Symbols Derived from Han Ideographs 756
CJK and Kangxi Radicals: U+2E80–U+2FD5 757
CJK Additions from HKSCS and GB 18030 758
CJK Strokes: U+31C0–U+31EF 758
Ideographic Symbols and Punctuation: U+16FE0–U+16FFF 759
18.2 Ideographic Description Characters 760
18.3 Bopomofo 764
18.4 Hiragana and Katakana 767
Hiragana: U+3040–U+309F 767
Katakana: U+30A0–U+30FF 767
Katakana Phonetic Extensions: U+31F0–U+31FF 768
Small Kana Extension: U+1B130-U+1B16F 768
Kana Supplement: U+1B000–U+1B0FF 768
Kana Extended-A: U+1B100–U+1B12F 768
Kana Extended-B: U+1AFF0-U+1AFFF 770
18.5 Halfwidth and Fullwidth Forms 773
18.6 Hangul 774
Hangul Jamo: U+1100–U+11FF 774
Hangul Jamo Extended-A: U+A960–U+A97F 775
Hangul Jamo Extended-B: U+D7B0–U+D7FF 775
Hangul Compatibility Jamo: U+3130–U+318F 775
Hangul Syllables: U+AC00–U+D7A3 776
18.7 Yi 778

18.8 Nüshu ... 781
18.9 Lisu ... 782
18.10 Miao ... 785
18.11 Tangut ... 787
Tangut: U+17000–U+187FF ... 787
Tangut Supplement: U+18D00–U+18D8F ... 787
Tangut Components: U+18800–U+18AFF ... 788
18.12 Khitan Small Script ... 790
19 Africa ... 793
19.1 Ethiopic ... 794
Ethiopic: U+1200–U+137F ... 794
Ethiopic Extensions ... 797
19.2 Osmanya ... 798
19.3 Tifinagh ... 799
19.4 N'Ko ... 802
19.5 Vai ... 807
19.6 Bamum ... 809
Bamum: U+A6A0–U+A6FF ... 809
Bamum Supplement: U+16800–U+16A3F ... 809
19.7 Bassa Vah ... 811
19.8 Mende Kikakui ... 812
19.9 Adlam ... 814
19.10 Medefaidrin ... 816
20 Americas ... 817
20.1 Cherokee ... 818
20.2 Canadian Aboriginal Syllabics ... 820
Unified Canadian Aboriginal Syllabics: U+1400–U+167F ... 820
Unified Canadian Aboriginal Syllabics Extended: U+18B0–U+18FF ... 821
Unified Canadian Aboriginal Syllabics Extended-A: U+11AB0–U+11ABF 821
20.3 Osage ... 822
20.4 Deseret ... 823
21 Notational Systems ... 827
21.1 Braille ... 828
21.2 Western Musical Symbols ... 830
21.3 Byzantine Musical Symbols ... 836
21.4 Znamenny Musical Notation ... 837
21.5 Ancient Greek Musical Notation ... 838
21.6 Duployan ... 841
Duployan: U+1BC00–U+1BC9F ... 841
Shorthand Format Controls: U+1BCA0–U+1BCAF ... 842

21.7 Sutton SignWriting 843
Sutton SignWriting: U+1D800–U+1DAAF 843
22 Symbols 845
22.1 Currency Symbols 847
Currency Symbols: U+20A0–U+20CF 847
22.2 Letterlike Symbols 851
Letterlike Symbols: U+2100–U+214F 851
Mathematical Alphanumeric Symbols: U+1D400–U+1D7FF 852
Mathematical Alphabets 853
Fonts Used for Mathematical Alphabets 856
Arabic Mathematical Alphabetic Symbols: U+1EE00–U+1EEFF 857
22.3 Numerals 859
Decimal Digits 859
Other Digits 861
Non-Decimal Radix Systems 864
Acrophonic Systems and Other Letter-based Numbers 865
Coptic Epact Numbers: U+102E0–U+102FF 866
Rumi Numeral Symbols: U+10E60–U+10E7E 867
Siyaq Numerical Notation Systems 868
CJK Numerals 869
Fractions 870
Common Indic Number Forms: U+A830–U+A83F 871
22.4 Superscript and Subscript Symbols 873
Superscripts and Subscripts: U+2070–U+209F 873
22.5 Mathematical Symbols 875
Mathematical Operators: U+2200–U+22FF 876
Supplements to Mathematical Symbols and Arrows 879
Supplemental Mathematical Operators: U+2A00–U+2AFF 879
Miscellaneous Mathematical Symbols-A: U+27C0–U+27EF 879
Miscellaneous Mathematical Symbols-B: U+2980–U+29FF 880
Miscellaneous Symbols and Arrows: U+2B00–U+2B7F 881
Arrows: U+2190–U+21FF 881
Supplemental Arrows 882
Standardized Variants of Mathematical Symbols 882
22.6 Invisible Mathematical Operators 883
22.7 Technical Symbols 884
Control Pictures: U+2400–U+243F 884
Miscellaneous Technical: U+2300–U+23FF 884
Optical Character Recognition: U+2440–U+245F 887
Symbols for Legacy Computing: U+1FB00-U+1FBFF 888

22.8 Geometrical Symbols 889
Box Drawing and Block Elements 889
Geometric Shapes: U+25A0–U+25FF 890
Geometric Shapes Extended: U+1F780–U+1F7FF 891
22.9 Miscellaneous Symbols 893
Miscellaneous Symbols and Pictographs 895
Emoticons: U+1F600–U+1F64F 898
Transport and Map Symbols: U+1F680–U+1F6FF 898
Dingbats: U+2700–U+27BF 898
Ornamental Dingbats: U+1F650–U+1F67F 899
Alchemical Symbols: U+1F700–U+1F77F 900
Mahjong Tiles: U+1F000–U+1F02F 900
Domino Tiles: U+1F030–U+1F09F 901
Playing Cards: U+1F0A0–U+1F0FF 901
Chess Symbols: U+1FA00–U+1FA6F 902
Yijing Hexagram Symbols: U+4DC0–U+4DFF 903
Tai Xuan Jing Symbols: U+1D300–U+1D356 903
Ancient Symbols: U+10190–U+101CF 904
Phaistos Disc Symbols: U+101D0–U+101FF 904
22.10 Enclosed and Square 906
Enclosed Alphanumerics: U+2460–U+24FF 908
Enclosed CJK Letters and Months: U+3200–U+32FF 908
CJK Compatibility: U+3300–U+33FF 909
Enclosed Alphanumeric Supplement: U+1F100–U+1F1FF 909
Enclosed Ideographic Supplement: U+1F200–U+1F2FF 910
23 Special Areas and Format Characters 913
23.1 Control Codes 914
Representing Control Sequences 914
Specification of Control Code Semantics 915
23.2 Layout Controls 917
Line and Word Breaking 917
Cursive Connection and Ligatures 919
Prepended Concatenation Marks 923
Combining Grapheme Joiner 924
Bidirectional Ordering Controls 926
Stateful Format Controls 927
23.3 Deprecated Format Characters 929
23.4 Variation Selectors 931
23.5 Private-Use Characters 934
Private Use Area: U+E000–U+F8FF 935
Supplementary Private Use Areas 936
23.6 Surrogates Area 937
23.7 Noncharacters 938

23.8 Specials ... 940
Byte Order Mark (BOM): U+FEFF ... 940
Specials: U+FFF0–U+FFF8 ... 942
Annotation Characters: U+FFF9–U+FFFB ... 942
Replacement Characters: U+FFFC–U+FFFD ... 944
23.9 Tag Characters ... 945
Tag Characters: U+E0000–U+E007F ... 945
Deprecated Use for Language Tagging ... 945
24 About the Code Charts ... 947
24.1 Character Names List ... 948
Images in the Code Charts and Character Lists ... 948
Special Characters and Code Points ... 950
Character Names ... 952
Informative Aliases ... 952
Normative Aliases ... 953
Cross References ... 954
Information About Languages ... 955
Case Mappings ... 956
Decompositions ... 956
Standardized Variation Sequences ... 957
Positional Forms ... 959
Block Headers ... 959
Subheads ... 960
24.2 CJK and Other Ideographs ... 961
CJK Unified Ideographs ... 961
Compatibility Ideographs ... 963
Tangut Ideographs ... 963
24.3 Hangul Syllables ... 965
A Notational Conventions ... 967
A.1 Typographic Conventions ... 968
Code Points ... 968
Character Names ... 968
Character Blocks ... 968
Sequences ... 969
Properties and Property Values ... 969
Miscellaneous ... 969
Operators ... 970
A.2 Extended BNF ... 971
Character Classes ... 971
A.3 Rendering ... 974

B Unicode Publications and Resources ... 975
B.1 The Unicode Consortium ... 976
The Unicode Technical Committee ... 976
Other Activities ... 976
B.2 Unicode Publications ... 977
B.3 Other Unicode Online Resources ... 978
Unicode Online Resources ... 978
How to Contact the Unicode Consortium ... 980
C Relationship to ISO/IEC 10646 ... 981
C.1 History ... 982
C.2 Encoding Forms in ISO/IEC 10646 ... 990
Zero Extending ... 990
C.3 UTF-8 and UTF-16 ... 991
UTF-8 ... 991
UTF-16 ... 991
C.4 Synchronization of the Standards ... 992
C.5 Identification of Features for Unicode ... 993
C.6 Character Names ... 994
C.7 Character Functional Specifications ... 995
D Version History of the Standard ... 997
E Han Unification History ... 999
E.1 Development of the URO ... 1000
E.2 Continuing Research on Ideographs ... 1002
Ideographic Rapporteur Group ... 1002
Ideographic Research Group ... 1003
E.3 CJK Sources ... 1004
F Documentation of CJK Strokes ... 1005
I Index ... 1013